

## A common repressor pool results in indeterminacy of extrinsic noise

Michail Stamatakis, Rhys M. Adams, and Gábor Balázsi

Citation: *Chaos* **21**, 047523 (2011); doi: 10.1063/1.3658618

View online: <http://dx.doi.org/10.1063/1.3658618>

View Table of Contents: <http://chaos.aip.org/resource/1/CHAOEH/v21/i4>

Published by the [American Institute of Physics](#).

---

### Related Articles

Key role of time-delay and connection topology in shaping the dynamics of noisy genetic regulatory networks  
*Chaos* **21**, 047522 (2011)

Logical stochastic resonance with correlated internal and external noises in a synthetic biological logic block  
*Chaos* **21**, 047521 (2011)

Hierarchical genetic networks and noncoding RNAs  
*Chaos* **20**, 045112 (2010)

Cell differentiation modeled via a coupled two-switch regulatory network  
*Chaos* **20**, 045121 (2010)

Simple genomes, complex interactions: Epistasis in RNA virus  
*Chaos* **20**, 026106 (2010)

---

### Additional information on Chaos

Journal Homepage: <http://chaos.aip.org/>

Journal Information: [http://chaos.aip.org/about/about\\_the\\_journal](http://chaos.aip.org/about/about_the_journal)

Top downloads: [http://chaos.aip.org/features/most\\_downloaded](http://chaos.aip.org/features/most_downloaded)

Information for Authors: <http://chaos.aip.org/authors>

### ADVERTISEMENT



*Submit Now*

### Explore AIP's new open-access journal

- Article-level metrics now available
- Join the conversation! Rate & comment on articles

# A common repressor pool results in indeterminacy of extrinsic noise

Michail Stamatakis,<sup>1,a)</sup> Rhys M. Adams,<sup>2</sup> and Gábor Balázsi<sup>2,b)</sup>

<sup>1</sup>Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, USA

<sup>2</sup>Department of Systems Biology – Unit 950, The University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA

(Received 21 July 2011; accepted 17 October 2011; published online 29 December 2011)

For just over a decade, stochastic gene expression has been the focus of many experimental and theoretical studies. It is now widely accepted that noise in gene expression can be decomposed into extrinsic and intrinsic components, which have orthogonal contributions to the total noise. Intrinsic noise stems from the random occurrence of biochemical reactions and is inherent to gene expression. Extrinsic noise originates from fluctuations in the concentrations of regulatory components or random transitions in the cell's state and is imposed to the gene of interest by the intra- and extra-cellular environment. The basic assumption has been that extrinsic noise acts as a pure input on the gene of interest, which exerts no feedback on the extrinsic noise source. Thus, multiple copies of a gene would be uniformly influenced by an extrinsic noise source. Here, we report that this assumption falls short when multiple genes share a common pool of a regulatory molecule. Due to the competitive utilization of the molecules existing in this pool, genes are no longer uniformly influenced by the extrinsic noise source. Rather, they exert negative regulation on each other and thus extrinsic noise cannot be determined by the currently established method.

© 2011 American Institute of Physics. [doi:10.1063/1.3658618]

Biology has been an immense source of problems demanding quantitative understanding. Efforts to take up this challenge have fueled the emergence of novel areas in mathematics and physics, including theories of random walks,<sup>1,2</sup> synchronization<sup>3</sup> deterministic chaos,<sup>4,5</sup> biological networks,<sup>6,7</sup> and stochastic resonance.<sup>8</sup> Without exception, each of these fields has been confronted with the issue of *noise*, defined as random and unpredictable fluctuations in the variables or parameters of the system being investigated. A remarkable example is stochastic gene expression, which has been the focus of increasing number of experimental and theoretical studies over the last two decades,<sup>9–16</sup> and is defined as the random variation of messenger RNA or protein levels in time in a single cell, or across a population of genetically identical cells at a given time. It is generally accepted that the sources of gene expression noise can be either extrinsic or intrinsic, which give rise to two orthogonal noise components, in the sense that their variances add up to the total variance.<sup>17–19</sup> *Intrinsic noise* originates from randomly occurring biochemical reactions during transcription and translation and is inherent to gene expression. *Extrinsic noise* can be attributed to fluctuations in the concentrations of regulatory components or random transitions in the cell's state and is imposed to the gene of interest by the intra- and extra-cellular environment. The separation of total noise into these components assumes that extrinsic noise acts as a pure input on the gene of interest, which does not have any effect (feedback) on the extrinsic noise

source. Thus, multiple copies of a gene exposed to a common extrinsic noise source should be identically influenced by their environment. Here, we report that this assumption cannot hold when multiple genes share a common pool of a regulatory molecule, because association with one of the target genes reduces the number of regulators available for other target genes. Therefore, genes are no longer uniformly influenced by the extrinsic noise source and exert indirect negative regulatory effect on each other. Consequently, extrinsic noise cannot be determined by the currently established method. We exemplify this conundrum by calculating the intrinsic and extrinsic noise in a simple theoretical system inspired from the original papers that described a method for separately measuring these noise components.<sup>17,18</sup> After some approximations, we obtain negative extrinsic noise values that have no physical meaning, both from analytically solving the Master equation and from performing stochastic simulations. These findings bring under question two key assumptions made when separating noise into intrinsic and extrinsic components: the “pure input” and the “independent genes” assumption. We conclude by discussing the implications of these results and proposing alternative strategies for defining noise components.

## I. INTRODUCTION

The basis for separating total gene expression noise into intrinsic and extrinsic components is the assumption that gene expression process is analogous to a noisy machine (Figure 1) that receives a noisy signal and performs a transformation of that signal while generating its own noise. Thus,

<sup>a)</sup>Current address: Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716, USA.

<sup>b)</sup>Author to whom correspondence should be addressed. Electronic mail: gbalazsi@mdanderson.org.

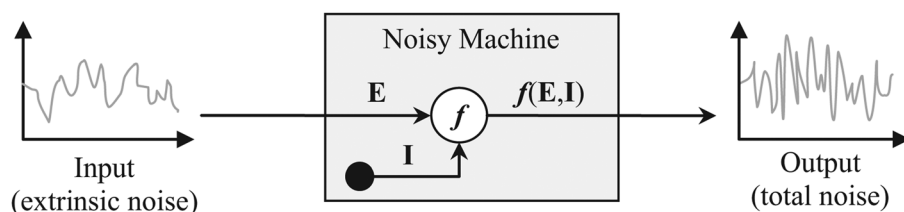


FIG. 1. A gene could be viewed as a noisy machine that accepts a random signal as input (extrinsic noise  $E$ ) and transforms it through an inherently stochastic process (intrinsic noise,  $I$ ), thereby generating a noisy output signal ( $f(E, I)$ ). Consequently, the latter contains two orthogonal noise components: extrinsic, due to the input noise, and intrinsic due to the noise produced within the “machine.” In our discussion, the output is the protein expression level:  $f(E, I) = P(E, I)$ .

the output of this machine will contain two noise components: the input (extrinsic) noise and the generated (intrinsic) noise. Analyzing the outputs of two identical and non-interacting such noisy machines exposed to the same input, one could safely assume that (i) the extrinsic components of these outputs would be identical since both machines receive the same input, (ii) the intrinsic components of the two outputs would be independent since the machines do not interact. Under these assumptions, the *extrinsic variables* can be defined as the noisy input variables to the noisy machines, while the *intrinsic variables* as the internal state variables of the noisy machines.

In the case of gene expression, these two assumptions can be stated as follows. First, the “pure input assumption” states that the gene of interest cannot exert any feedback on the extrinsic noise source. Therefore, extrinsic noise acts as a pure input to the process of gene expression. This assumption allows for the partitioning into extrinsic and intrinsic variables. In particular, Swain *et al.*<sup>17</sup> give examples of extrinsic variables being the state of the cell or the levels of regulatory components, and examples of intrinsic variables being the mRNA and protein contents for the gene of interest. Second, the “independent genes assumption” states that multiple identical copies of a gene do not influence one-another. This second assumption makes possible the calculation of the intrinsic and extrinsic noise components by simultaneously measuring the protein contents of two identical gene copies.<sup>17</sup>

Even though the decomposition of noise into extrinsic and intrinsic components has been a useful tool for analyzing stochasticity in gene expression, such an analysis is questionable unless both of the aforementioned assumptions hold true. In particular, this may not always be the case in biological systems: metabolic or signaling pathways, as well as synthetic gene networks often contain either explicit or implicit feedback loops.<sup>20,21</sup> The components of such feedback loops would contribute to their own extrinsic noise, which would invalidate the “pure input assumption”. Furthermore, autoregulation,<sup>22</sup> namely, a gene exerting feedback to its own expression would also invalidate the “independent genes assumption.” Likewise, two identical copies of a gene that share a common pool of a low-abundance regulatory species<sup>23</sup> could exert indirect negative regulatory effects on each other, due to the competitive utilization of the shared regulator pool. Such an effect is plausible in naturally occurring genetic networks, such as the *lac* operon in which the LacI repressor concentration is roughly 10 copies per cell.<sup>24</sup>

The aim of this paper is to examine a case involving a shared regulator pool and demonstrate that these two assumptions are violated; thus, even partitioning the overall

noise into extrinsic and intrinsic components becomes problematic. We will apply mathematical modeling to a system similar to the one introduced to experimentally measure the intrinsic and extrinsic noise,<sup>18</sup> which comprises of two reporter-gene variants under the influence of identical promoters repressed by LacI. We will show that the “pure input assumption” does not hold, because the state of each reporter gene’s promoter (repressed or not) affects the free repressor content. Likewise, the “independent genes assumption” fails, because if one promoter is in the repressed state, then this prevents one repressor molecule from repressing the other gene. Thus, the competitive utilization of a common repressor pool results in mutual indirect negative regulatory interactions, and subsequently in negatively correlated reporter protein contents. If one tries to calculate the extrinsic noise for this system, negative values with no physical meaning may be encountered.

The rest of the paper is organized as follows. We will first briefly overview the concepts of intrinsic and extrinsic noise. Second, we will propose a network of reactions that captures the salient features of the underlying biochemical interactions. Next, we will derive a stochastic model for the transitions of the two operators to the repressed and unrepressed states. We will then show that the “pure input assumption” and the “independent genes assumption” can only hold in the limit of an infinitely large repressor pool. Otherwise, meaningless negative extrinsic noise values may be encountered. Finally, we conclude by discussing when it is meaningful to consider partitioning the noise into extrinsic and intrinsic components and what alternatives may exist for properly defining them.

## II. INTRINSIC AND EXTRINSIC NOISE: AN OVERVIEW

Much before the first gene expression measurements in single cells, the problem of information transfer through noisy nonlinear systems has confronted neuroscience for decades. Noise was originally considered a universal hindrance for information transfer in neurons,<sup>25</sup> but this view had to be revised when noise-optimized weak signal detection and transfer were demonstrated experimentally<sup>26</sup> and explained by theory<sup>27,28</sup> in neuronal systems. While externally added noise showed capable of optimizing signal transfer in biological neurons, a key question that remained open was whether noise internal to biological neurons, such as the noise originating from random ion channel dynamics<sup>29</sup> could do the same. Theoretical work suggested that this may be possible,<sup>30,31</sup> but a rigorous experimental demonstration is still lacking. Posing this question treated neurons as noisy machines that could add

their own intrinsic noise to the extrinsic noise already present in the incoming signal (Figure 1).

The issue of intrinsic and extrinsic noise arose again once gene expression noise became directly measurable in living cells. By using a chromosomally integrated fluorescent reporter, van Oudenaarden and colleagues studied the sources of random intrinsic gene expression variations in genetically identical *Bacillus subtilis* cells.<sup>32</sup> Later that year, Elowitz and colleagues reported on the intrinsic and extrinsic components of gene expression noise in *Escherichia coli* using a cleverly designed *double-reporter system*<sup>18</sup> that involved two fluorescent reporters located at equivalent chromosomal loci, following theoretical predictions by Swain *et al.*<sup>17</sup> As sources of intrinsic noise, they propose random transcription and translation initiation events, as well as degradation of transcripts and proteins. Extrinsic noise, on the other hand originates from cell-wide fluctuations in the number of ribosomes, polymerases, regulators, cell cycle stage, etc. Since these original studies in bacteria, the double-reporter system has been successfully adapted to other organisms.<sup>33–35</sup> Besides the double-reporter system, other methods proposed for estimating the intrinsic and extrinsic components of gene expression noise have relied on total noise measurements for various copy numbers of a reporter gene<sup>36,37</sup> or mutations altering biological parameters relevant for intrinsic noise only (such as the promoter sequence, and the rates of transcription and translation).<sup>32,38,39</sup>

Below we summarize the essence of the theory<sup>17</sup> underlying the use of the double-reporter methodology<sup>18</sup> to separate the intrinsic and extrinsic components of gene expression noise.

Let us consider a biological process for which the observable of interest is the expression level of a protein. Swain *et al.*<sup>17</sup> propose that this level is a function  $P(\mathbf{E}, \mathbf{I})$  of extrinsic and intrinsic variables  $\mathbf{E}$  and  $\mathbf{I}$ . As the variables  $\mathbf{E}$  and  $\mathbf{I}$  fluctuate, so does the observed (measured) expression level  $P(\mathbf{E}, \mathbf{I})$ , and the total variance of the observable will be

$$\langle (P - \langle P \rangle)^2 \rangle = \iint (P(\mathbf{E}, \mathbf{I}) - \langle P \rangle)^2 \rho(\mathbf{E}, \mathbf{I}) d\mathbf{E} d\mathbf{I}, \quad (1)$$

where  $\rho(\mathbf{E}, \mathbf{I})$  is the joint probability density of  $\mathbf{E}$ ,  $\mathbf{I}$ , and:

$$\langle P \rangle = \int P(\mathbf{E}, \mathbf{I}) \rho(\mathbf{E}, \mathbf{I}) d\mathbf{E} d\mathbf{I}. \quad (2)$$

From Bayes' theorem, the joint probability  $\rho(\mathbf{E}, \mathbf{I})$  can be expressed as

$$\rho(\mathbf{E}, \mathbf{I}) = \rho(\mathbf{I}|\mathbf{E}) \cdot \rho(\mathbf{E}). \quad (3)$$

Thus, from Eq. (1) by adding and subtracting a term  $\int (\int P(\mathbf{E}, \mathbf{I}) \rho(\mathbf{I}|\mathbf{E}) d\mathbf{I})^2 \rho(\mathbf{E}) d\mathbf{E}$  and performing some algebraic manipulations they obtain

$$\langle (P - \langle P \rangle)^2 \rangle = \underbrace{\int \left( \int P(\mathbf{E}, \mathbf{I})^2 \rho(\mathbf{I}|\mathbf{E}) d\mathbf{I} - \left( \int P(\mathbf{E}, \mathbf{I}) \rho(\mathbf{I}|\mathbf{E}) d\mathbf{I} \right)^2 \right) \rho(\mathbf{E}) d\mathbf{E}}_{\text{intrinsic variance}} + \underbrace{\int \left( \int P(\mathbf{E}, \mathbf{I}) \rho(\mathbf{I}|\mathbf{E}) d\mathbf{I} \right)^2 \rho(\mathbf{E}) d\mathbf{E} - \langle P \rangle^2}_{\text{extrinsic variance}}. \quad (4)$$

In order to calculate the extrinsic and intrinsic noise contributions just noted, Swain *et al.*<sup>17</sup> suggest using two identical expression systems (noisy machines) for which both reporter genes are exposed to the same extrinsic variables while having individual intrinsic variables with identical probability densities

$$\mathbf{E}_1 = \mathbf{E}_2 = \mathbf{E}, \quad (5)$$

$$\rho(\mathbf{I}_1|\mathbf{E}) = \rho(\mathbf{I}_2|\mathbf{E}). \quad (6)$$

This allows the calculation of the coupling term that was added and subtracted

$$\begin{aligned} \int \left( \int P(\mathbf{E}, \mathbf{I}) \rho(\mathbf{I}|\mathbf{E}) d\mathbf{I} \right)^2 \rho(\mathbf{E}) d\mathbf{E} &= \int \left( \int P(\mathbf{E}_1, \mathbf{I}_1) \rho(\mathbf{I}_1|\mathbf{E}_1) d\mathbf{I}_1 \right) \left( \int P(\mathbf{E}_2, \mathbf{I}_2) \rho(\mathbf{I}_2|\mathbf{E}_2) d\mathbf{I}_2 \right) \rho(\mathbf{E}_1, \mathbf{E}_2) \delta(\mathbf{E}_1 - \mathbf{E}_2) d\mathbf{E}_1 d\mathbf{E}_2 \\ &= \int \int P(\mathbf{E}_1, \mathbf{I}_1) P(\mathbf{E}_2, \mathbf{I}_2) \delta(\mathbf{E}_1 - \mathbf{E}_2) \rho(\mathbf{E}_1, \mathbf{E}_2, \mathbf{I}_1, \mathbf{I}_2) d\mathbf{I}_1 d\mathbf{I}_2 d\mathbf{E}_1 d\mathbf{E}_2. \end{aligned} \quad (7)$$

Let  $P(\mathbf{E}_1, \mathbf{I}_1) = c$  be the amount of cyan fluorescent protein (CFP) in the cell, and  $P(\mathbf{E}_2, \mathbf{I}_2) = y$  the amount of yellow fluorescent protein (YFP). These amounts can be inferred from the corresponding fluorescence values measured for individual cells by flow cytometry or fluorescence microscopy. Equations (5) and (6) hold true, assuming that both reporter genes are under the influence of identical promoters and integrated into equivalent chromosomal loci. Consequently, according to Eq. (7), the coupling term is simply the cross-correlation between  $c$  and  $y$  measured in the same cell

$$\int \left( \int P(\mathbf{E}, \mathbf{I}) \rho(\mathbf{I}|\mathbf{E}) d\mathbf{I} \right)^2 \rho(\mathbf{E}) d\mathbf{E} = \langle cy \rangle, \quad (8)$$

where the angled brackets indicate averaging over the cell population. Furthermore, from the equivalence of the two reporter genes

$$\begin{aligned} \iint P(\mathbf{E}, \mathbf{I})^2 \rho(\mathbf{E}, \mathbf{I}) d\mathbf{E} d\mathbf{I} &= \langle c^2 \rangle = \langle y^2 \rangle, \\ \iint P(\mathbf{E}, \mathbf{I}) \rho(\mathbf{E}, \mathbf{I}) d\mathbf{E} d\mathbf{I} &= \langle c \rangle = \langle y \rangle, \end{aligned} \quad (9)$$



and by substituting Eqs. (9) back to Eq. (4)

$$\langle (P - \langle P \rangle)^2 \rangle = \underbrace{\langle c^2 \rangle - \langle cy \rangle}_{\text{intrinsic variance}} + \underbrace{\langle cy \rangle - \langle c \rangle^2}_{\text{extrinsic variance}}. \quad (10)$$

In the above equation, the intrinsic variance can be further simplified as follows:

$$\langle c^2 \rangle - \langle cy \rangle = \frac{1}{2} \cdot (\langle c^2 \rangle + \langle y^2 \rangle) - \langle cy \rangle = \frac{1}{2} \cdot \langle (c - y)^2 \rangle. \quad (11)$$

This procedure allows for the estimation of intrinsic and extrinsic noise components measured as the coefficient of variation (CV; standard deviation divided by the mean) based on the formula

$$\eta_{tot}^2 = \frac{\langle (P - \langle P \rangle)^2 \rangle}{\langle P \rangle^2} = \frac{\langle (c - y)^2 \rangle}{2\langle c \rangle \langle y \rangle} + \frac{\langle cy \rangle - \langle c \rangle \langle y \rangle}{\langle c \rangle \langle y \rangle} = \frac{\langle c^2 + y^2 \rangle - 2\langle c \rangle \langle y \rangle}{2\langle c \rangle \langle y \rangle}. \quad (12)$$

Note that the first equality of Eq. (7) depends on the assumption that the expression systems are always exposed to identical external variables in Eq. (5). Is this always true or can there be exceptions? In the following, we will discuss a case when this assumption is not satisfied and show how this can give unphysical results when calculating the intrinsic and extrinsic noise components.

### III. MATHEMATICAL MODELING AND STOCHASTIC SIMULATIONS

#### A. Competitive repressor utilization

##### 1. Formulation of the model

As an example for unequal extrinsic variables, we consider a two-promoter system regulated by the LacI repressor, similar to the system built to experimentally determine the intrinsic and extrinsic components of gene expression noise.<sup>18</sup> The molecular mechanisms included in our model are depicted in Figure 2; species notation is shown in Table I, kinetic constants are shown in Table II, and the reaction network is summarized in Table III. In constructing the

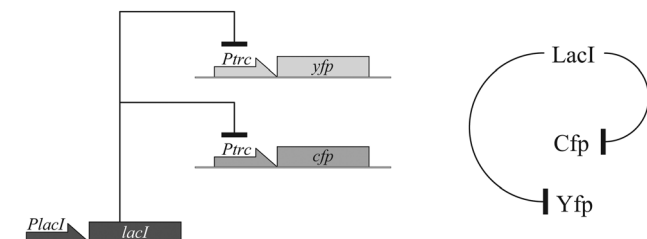


FIG. 2. Schematic representation of the interactions taken into account in the two promoter-reporter system. For species notation see Table I.

TABLE I. Symbols used for the species.

Symbol	Species denoted
Lac	LacI repressor
O <sub>yfp</sub>	yfp lacO operator
O <sub>cfp</sub>	cfp lacO operator
R <sub>yfp</sub>	yfp m-RNA
R <sub>cfp</sub>	cfp m-RNA
Yfp	Yfp
Cfp	Cfp
∅	Generic source or sink

TABLE II. Kinetic constants (deterministic and stochastic).

Deterministic	Value	Units	Relation to stochastic <sup>a</sup>
$\bar{k}_{Lac}$	0.23	nM·min <sup>-1</sup>	$k_{Lac} = \bar{k}_{Lac} \cdot N_A \cdot V_{E.coli}$
$\bar{k}_r$	15	nM <sup>-1</sup> ·min <sup>-1</sup>	$k_r = \frac{\bar{k}_r}{N_A \cdot V_{E.coli}}$
$\bar{k}_{-r}$	50	min <sup>-1</sup>	$k_{-r} = \bar{k}_{-r}$
$\bar{k}_m$	50	min <sup>-1</sup>	$k_m = \bar{k}_m$
$\bar{k}_p$	50	min <sup>-1</sup>	$k_p = \bar{k}_p$
$\bar{\lambda}_{lac}$	50	min <sup>-1</sup>	$\lambda_{lac} = \bar{\lambda}_{lac}$
$\bar{\lambda}_m$	50	min <sup>-1</sup>	$\lambda_m = \bar{\lambda}_m$
$\bar{\lambda}_p$	10 <sup>-3</sup>	min <sup>-1</sup>	$\lambda_p = \bar{\lambda}_p$

<sup>a</sup>Avogadro's number:  $N_A = 6.0221367 \cdot 10^{14}$  nmol<sup>-1</sup> and *E. coli* volume was taken  $V_{E.coli} = 8 \cdot 10^{-16}$  L

reaction network, for simplicity we assume constitutive LacI production, single *lacO* operator sites per promoter,<sup>40</sup> abundant RNA polymerases and ribosomes and 1<sup>st</sup> order degradation reactions. We ignore the effects of cell division and DNA duplication as we want to concentrate on the effects of competitive utilization of the LacI repressor pool. Thus, the stochastic model for this system is the chemical Master equation

TABLE III. Reaction network and propensity functions.

Reaction	Propensity function <sup>a,b</sup>
(1) $\emptyset \xrightarrow{k_{Lac}} Lac$	$\alpha_1 = k_{Lac}$
(2) $O_{yfp} + Lac \xrightarrow{k_r} O_{yfp}Lac$	$\alpha_2 = k_r \cdot O_{yfp} \cdot Lac$
(3) $O_{yfp}Lac \xrightarrow{k_{-r}} O_{yfp} + Lac$	$\alpha_3 = k_{-r} \cdot O_{yfp}Lac$
(4) $O_{yfp} \xrightarrow{k_m} O_{yfp} + R_{yfp}$	$\alpha_4 = k_m \cdot O_{yfp}$
(5) $R_{yfp} \xrightarrow{k_p} R_{yfp} + Yfp$	$\alpha_5 = k_p \cdot R_{yfp}$
(6) $O_{cfp} + Lac \xrightarrow{k_r} O_{cfp}Lac$	$\alpha_6 = k_r \cdot O_{cfp} \cdot Lac$
(7) $O_{cfp}Lac \xrightarrow{k_{-r}} O_{cfp} + Lac$	$\alpha_7 = k_{-r} \cdot O_{cfp}Lac$
(8) $O_{cfp} \xrightarrow{k_m} O_{cfp} + R_{cfp}$	$\alpha_8 = k_m \cdot O_{cfp}$
(9) $R_{cfp} \xrightarrow{k_p} R_{cfp} + Cfp$	$\alpha_9 = k_p \cdot R_{cfp}$
(10) $Lac \xrightarrow{\lambda_{Lac}} \emptyset$	$\alpha_{10} = \lambda_{Lac} \cdot Lac$
(11) $R_{yfp} \xrightarrow{\lambda_p} \emptyset$	$\alpha_{11} = \lambda_p \cdot R_{yfp}$
(12) $Yfp \xrightarrow{\lambda_p} \emptyset$	$\alpha_{12} = \lambda_p \cdot Yfp$
(13) $R_{cfp} \xrightarrow{\lambda_m} \emptyset$	$\alpha_{13} = \lambda_m \cdot R_{cfp}$
(14) $Cfp \xrightarrow{\lambda_p} \emptyset$	$\alpha_{14} = \lambda_p \cdot Cfp$

<sup>a</sup>Variables without brackets denote number of molecules of the corresponding species.

<sup>b</sup>All propensity functions have units of min<sup>-1</sup>.

$$\frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t)}{\partial t} = \sum_{j=1}^m [\alpha_j(\mathbf{x} - \mathbf{v}_j) \cdot P(\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t) - \alpha_j(\mathbf{x}) \cdot P(\mathbf{x}, t | \mathbf{x}_0, t)] \quad (13)$$

which can be simulated with the Gillespie algorithm.<sup>41,42</sup>

The state vector containing numbers of molecules for each species is  $\mathbf{x} = \{\text{Lac}, \text{O}_{Yfp}, \text{O}_{Yfp}\text{Lac}, \text{O}_{Cfp}, \text{O}_{Cfp}\text{Lac}, \text{R}_{Yfp}, \text{Yfp}, \text{R}_{Cfp}, \text{Cfp}\}$ , and we have  $n = 9$  species participating in  $m = 14$  reactions. The reactions' propensity functions,  $\alpha_j(\mathbf{x}) = c_j \cdot h_j(\mathbf{x})$ ,  $j = 1, 2, \dots, m$ , are presented in Table III. The

vectors  $\mathbf{v}_j$  denote the change in the number of molecules for each species, e.g., for reaction (6),  $\mathbf{v}_6 = \{-1, 0, 0, -1, 1, 0, 0, 0, 0\}$ .

Let us now define the step operator as follows:<sup>43</sup>

$$\begin{aligned} E_m f(k, l, m, n, \dots) &= f(k, l, m + 1, n, \dots), \\ E_m^{-1} f(k, l, m, n, \dots) &= f(k, l, m - 1, n, \dots). \end{aligned} \quad (14)$$

A useful property of the step operator is the following:

$$\sum_{m=0}^{N-1} g(k, l, m, n, \dots) \cdot E_m f(k, l, m, n, \dots) = \sum_{m=1}^N E_m^{-1} g(k, l, m, n, \dots) \cdot f(k, l, m, n, \dots). \quad (15)$$

Then the Master equation can be written as

$$\begin{aligned} \frac{\partial P}{\partial t} &= (E_{Lac}^{-1} - 1)(k_{Lac} \cdot P) + (E_{Lac} - 1)(\lambda_{Lac} \cdot Lac \cdot P) + \\ &\quad (E_{O_{Yfp}} E_{Lac} E_{O_{Yfp}Lac}^{-1} - 1)(k_r \cdot O_{Yfp} \cdot Lac \cdot P) + (E_{O_{Yfp}}^{-1} E_{Lac}^{-1} E_{O_{Yfp}Lac} - 1)(k_{-r} \cdot O_{Yfp}Lac \cdot P) + \\ &\quad (E_{R_{Yfp}}^{-1} - 1)(k_m \cdot O_{Yfp} \cdot P) + (E_{R_{Yfp}} - 1)(\lambda_m \cdot R_{Yfp} \cdot P) + \\ &\quad (E_{Yfp}^{-1} - 1)(k_p \cdot R_{Yfp} \cdot P) + (E_{Yfp} - 1)(\lambda_p \cdot Yfp \cdot P) + \\ &\quad (E_{O_{Cfp}} E_{Lac} E_{O_{Cfp}Lac}^{-1} - 1)(k_r \cdot O_{Cfp} \cdot Lac \cdot P) + (E_{O_{Cfp}}^{-1} E_{Lac}^{-1} E_{O_{Cfp}Lac} - 1)(k_{-r} \cdot O_{Cfp}Lac \cdot P) + \\ &\quad (E_{R_{Cfp}}^{-1} - 1)(k_m \cdot O_{Cfp} \cdot P) + (E_{R_{Cfp}} - 1)(\lambda_m \cdot R_{Cfp} \cdot P) + \\ &\quad (E_{Cfp}^{-1} - 1)(k_p \cdot R_{Cfp} \cdot P) + (E_{Cfp} - 1)(\lambda_p \cdot Cfp \cdot P), \end{aligned} \quad (16)$$

where  $P = P(\text{Lac}, O_{Yfp}, O_{Yfp}\text{Lac}, O_{Cfp}, O_{Cfp}\text{Lac}, R_{Yfp}, Yfp, R_{Cfp}, Cfp; t)$ .

Next, we will apply a set of simplifying transformations to the Master equation, recasting it into a form that allows us to focus on the source of the problem defining extrinsic noise: the competition between Lac operator sites in the two promoters for LacI repressor molecules.

If we apply the summing operator  $\sum_{R_{Yfp} \geq 0} \sum_{Yfp \geq 0} \sum_{R_{Cfp} \geq 0} \sum_{Cfp \geq 0}$  to Eq. (16) using property (15) and taking into account boundary conditions such as  $E_{R_{Yfp}}^{-1}(k_m \cdot O_{Yfp} \cdot P) = 0$  for  $R_{Yfp} = 0$

$$\begin{aligned} \frac{\partial P}{\partial t} &= (E_{Lac}^{-1} - 1)(k_{Lac} \cdot P) + (E_{Lac} - 1)(\lambda_{Lac} \cdot Lac \cdot P) + \\ &\quad (E_{O_{Yfp}} E_{Lac} E_{O_{Yfp}Lac}^{-1} - 1)(k_r \cdot O_{Yfp} \cdot Lac \cdot P) + (E_{O_{Yfp}}^{-1} E_{Lac}^{-1} E_{O_{Yfp}Lac} - 1)(k_{-r} \cdot O_{Yfp}Lac \cdot P) + \\ &\quad (E_{O_{Cfp}} E_{Lac} E_{O_{Cfp}Lac}^{-1} - 1)(k_r \cdot O_{Cfp} \cdot Lac \cdot P) + (E_{O_{Cfp}}^{-1} E_{Lac}^{-1} E_{O_{Cfp}Lac} - 1)(k_{-r} \cdot O_{Cfp}Lac \cdot P), \end{aligned} \quad (17)$$

where  $P = P(\text{Lac}, O_{Yfp}, O_{Yfp}\text{Lac}, O_{Cfp}, O_{Cfp}\text{Lac}; t)$  is the marginal probability that allows us to focus on *cfp* and *yfp* promoter dynamics. We note that: (i)  $O_{Yfp}$  and  $O_{Cfp}$  are zero or unity, (ii) conservation of the operator sites holds

$$\begin{aligned} O_{Yfp} + O_{Yfp}\text{Lac} &= O_{Yfp,T}, \\ O_{Cfp} + O_{Cfp}\text{Lac} &= O_{Cfp,T}, \end{aligned} \quad (18)$$

and (iii) the operator fluctuations between the free and the repressed state do not change the overall repressor content, defined as

$$Lac_T = Lac + O_{Yfp}\text{Lac} + O_{Cfp}\text{Lac}. \quad (19)$$

We therefore only need three variables namely  $Lac_T$ ,  $O_{Yfp}$ , and  $O_{Cfp}$  to refer to the marginal probability  $P(\text{Lac}, O_{Yfp}, O_{Yfp}\text{Lac}, O_{Cfp}, O_{Cfp}\text{Lac}; t)$ . Thus, we define a new probability  $P$  as

$$\begin{aligned} \hat{P}(Lac_T, O_{Yfp}, O_{Cfp}; t) &= P(Lac_T - (O_{Yfp,T} - O_{Yfp}) \\ &\quad - (O_{Cfp,T} - O_{Cfp}), O_{Yfp}, O_{Yfp,T} - O_{Yfp}, O_{Cfp}, O_{Cfp,T} - O_{Cfp}; t), \end{aligned} \quad (20)$$

and we recast the Master equation as follows:

$$\begin{aligned} \frac{\partial \hat{P}}{\partial t} = & \left( E_{Lac_T}^{-1} - 1 \right) \left( k_{Lac} \cdot \hat{P} \right) + \left( E_{Lac_T} - 1 \right) \left( \lambda_{Lac} \cdot (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P} \right) + \\ & \left( E_{O_{Yfp}} - 1 \right) \left( k_r \cdot O_{Yfp} \cdot (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P} \right) + \left( E_{O_{Yfp}}^{-1} - 1 \right) \left( k_{-r} \cdot (O_{Yfp,T} - O_{Yfp}) \cdot \hat{P} \right) + \\ & \left( E_{O_{Cfp}} - 1 \right) \left( k_r \cdot O_{Cfp} \cdot (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P} \right) + \left( E_{O_{Cfp}}^{-1} - 1 \right) \left( k_{-r} \cdot (O_{Cfp,T} - O_{Cfp}) \cdot \hat{P} \right), \end{aligned} \quad (21)$$

If we now apply the summing operator  $\sum_{O_{Yfp}=0}^1 \sum_{O_{Cfp}=0}^1$  to Eq. (21) we get an equation describing the probability of total repressor molecules

$$\frac{\partial \hat{P}(Lac_T; t)}{\partial t} = \left( E_{Lac_T}^{-1} - 1 \right) \left( k_{Lac} \cdot \hat{P}(Lac_T; t) \right) + \left( E_{Lac_T} - 1 \right) \left( \lambda_{Lac} \cdot F(Lac_T) \cdot \hat{P}(Lac_T; t) \right) \quad (22)$$

where

$$F(Lac_T) = \sum_{O_{Yfp}=0}^1 \sum_{O_{Cfp}=0}^1 \left\{ (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P}(O_{Yfp}, O_{Cfp} | Lac_T; t) \right\}.$$

## 2. Fast-operator-fluctuations approximation

Next, based on the separation of time scales,<sup>44</sup> we establish equations separately describing the probability of total repressor number and target promoter states. For this, let us assume fast operator fluctuations

$$\begin{aligned} k_r &= \frac{\kappa_r}{\varepsilon}, \\ k_{-r} &= \frac{\kappa_{-r}}{\varepsilon}. \end{aligned} \quad (23)$$

This means that, due to fast repressor-operator binding, the probability of *cfp* and *yfp* operator states reaches steady state fast enough to be separable from the probability of having a

certain level of total LacI repressor. Therefore, we introduce expansions for the time-dependent probabilities  $P_0$  describing the operator states conditioned on total LacI and  $P_1$  describing the distribution of total LacI repressor level

$$\begin{aligned} \hat{P}(O_{Yfp}, O_{Cfp} | Lac_T; t) &= \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) + \varepsilon \\ &\quad \cdot \hat{P}_1(O_{Yfp}, O_{Cfp} | Lac_T; t) + \dots, \end{aligned} \quad (24)$$

$$\hat{P}(Lac_T; t) = \hat{P}_0(Lac_T; t) + \varepsilon \cdot \hat{P}_1(Lac_T; t) + \dots \quad (25)$$

From the normalization conditions, it can be derived that

$$\begin{array}{lll} \mathcal{O}(1) & \mathcal{O}(\varepsilon) & \text{HigherOrders} \\ 1 = \sum_{O_{Yfp}=0}^1 \sum_{O_{Cfp}=0}^1 \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) & 0 = \sum_{O_{Yfp}=0}^1 \sum_{O_{Cfp}=0}^1 \hat{P}_1(O_{Yfp}, O_{Cfp} | Lac_T; t) & \dots \end{array} \quad (26)$$

$$1 = \sum_{Lac \geq 0} \hat{P}_0(Lac; t) \quad 0 = \sum_{Lac \geq 0} \hat{P}_1(Lac; t) \quad \dots \quad (27)$$

Also

$$\begin{aligned} \hat{P}(Lac_T, O_{Yfp}, O_{Cfp}; t) &= \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \hat{P}_0(Lac_T; t) + \\ &\varepsilon \cdot \left[ \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \hat{P}_1(Lac_T; t) + \hat{P}_1(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \hat{P}_0(Lac_T; t) \right] + \\ &\varepsilon^2 \cdot \left[ \hat{P}_1(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \hat{P}_1(Lac_T; t) + \dots \right] + \dots, \end{aligned} \quad (28)$$

and

$$\begin{aligned} \frac{\partial}{\partial t} \hat{P}(Lac_T, O_{Yfp}, O_{Cfp}; t) &= \frac{\partial}{\partial t} \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \hat{P}_0(Lac_T; t) + \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \frac{\partial}{\partial t} \hat{P}_0(Lac_T; t) + \\ &\varepsilon \cdot \left[ \frac{\partial}{\partial t} \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \hat{P}_1(Lac_T; t) + \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \frac{\partial}{\partial t} \hat{P}_1(Lac_T; t) + \right. \\ &\left. \frac{\partial}{\partial t} \hat{P}_1(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \hat{P}_0(Lac_T; t) + \hat{P}_1(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \frac{\partial}{\partial t} \hat{P}_0(Lac_T; t) \right] + \dots \end{aligned} \quad (29)$$

Now let us go back to Eq. (21), in which we substitute expressions (28) and (23), and derive the expressions of order  $\epsilon^{-1}$  and 1

$$0 = (E_{O_{Yfp}} - 1) \left( \kappa_r \cdot O_{Yfp} \cdot (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P}_0 \right) + (E_{O_{Yfp}}^{-1} - 1) \left( \kappa_{-r} \cdot (O_{Yfp,T} - O_{Yfp}) \cdot \hat{P}_0 \right) + (E_{O_{Cfp}} - 1) \left( \kappa_r \cdot O_{Cfp} \cdot (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P}_0 \right) + (E_{O_{Cfp}}^{-1} - 1) \left( \kappa_{-r} \cdot (O_{Cfp,T} - O_{Cfp}) \cdot \hat{P}_0 \right). \quad (30)$$

Note that  $\hat{P}_0 = \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \cdot \hat{P}_0(Lac_T; t)$  and that we have no step operator  $E_{Lac}$  in Eq. (30). Therefore, we can write

$$0 = (E_{O_{Yfp}} - 1) \left( \kappa_r \cdot O_{Yfp} \cdot (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \right) + (E_{O_{Yfp}}^{-1} - 1) \left( \kappa_{-r} \cdot (O_{Yfp,T} - O_{Yfp}) \cdot \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \right) + (E_{O_{Cfp}} - 1) \left( \kappa_r \cdot O_{Cfp} \cdot (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \right) + (E_{O_{Cfp}}^{-1} - 1) \left( \kappa_{-r} \cdot (O_{Cfp,T} - O_{Cfp}) \cdot \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \right). \quad (31)$$

Furthermore, from Eq. (21)

$$\frac{\partial \hat{P}_0(Lac_T; t)}{\partial t} = (E_{Lac}^{-1} - 1) \left( k_{Lac} \cdot \hat{P}_0(Lac_T; t) \right) + (E_{Lac} - 1) \left( \lambda_{Lac} \cdot F_0(Lac_T) \cdot \hat{P}_0(Lac_T; t) \right) \quad (32)$$

where

$$F_0(Lac_T) = \sum_{O_{Yfp}=0}^1 \sum_{O_{Cfp}=0}^1 \left\{ (Lac_T - (O_{Yfp,T} - O_{Yfp}) - (O_{Cfp,T} - O_{Cfp})) \cdot \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \right\}.$$

Equation (32) is easy to solve, thereby finding the probabilities for the states of both operators given  $Lac_T$ . Then, one can substitute these probabilities into the expression for  $F_0$  in the stationary form of Eqs. (32) finding the stationary probability distribution for  $Lac_T$ .

### 3. Solution of the Master equation for *LacI* and its target operators

In the following, we will obtain exact expressions for the probabilities of having either *LacI*-bound or *LacI*-free promoter states for both *cfp* and *yfp* at a predefined, constant *LacI* level (which is valid for the quasi steady-state approxi-

mation that we applied). These expressions will allow later the calculation of intrinsic and extrinsic noise in terms of these promoter states. Let us therefore solve equation (26) at stationary conditions given  $Lac_T$  fixed. We will set

$$p_{O_{Yfp} O_{Cfp}} = \hat{P}_0(O_{Yfp}, O_{Cfp} | Lac_T; t) \quad (33)$$

$$q = Lac_T.$$

Therefore,  $p_{10}$  will denote the probability of finding  $O_{Yfp}$  in the free state, and  $O_{Cfp}$  in the repressed state. Since we have two operators each of which can be found in two different states, Eq.(31) is a  $4 \times 4$  linear system

$$\begin{bmatrix} -2 \cdot q \cdot \kappa_r & \kappa_{-r} & \kappa_{-r} & 0 \\ q \cdot \kappa_r & -\kappa_{-r} - (q-1) \cdot \kappa_r & 0 & \kappa_{-r} \\ q \cdot \kappa_r & 0 & -\kappa_{-r} - (q-1) \cdot \kappa_r & \kappa_{-r} \\ 0 & (q-1) \cdot \kappa_r & (q-1) \cdot \kappa_r & -2 \cdot \kappa_{-r} \end{bmatrix} \cdot \begin{bmatrix} p_{11} \\ p_{01} \\ p_{10} \\ p_{00} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (34)$$

whose solution can be calculated (using the normalization condition  $p_{00} + p_{01} + p_{10} + p_{11} = 1$ )

$$p_{00} = \frac{q \cdot (q-1) \cdot K_r^2}{q \cdot (q-1) \cdot K_r^2 + 2 \cdot q \cdot K_r + 1}$$

$$p_{01} = p_{10} = \frac{q \cdot K_r}{q \cdot (q-1) \cdot K_r^2 + 2 \cdot q \cdot K_r + 1}$$

$$p_{11} = \frac{1}{q \cdot (q-1) \cdot K_r^2 + 2 \cdot q \cdot K_r + 1} \quad (35)$$

where:

$$K_r = \frac{\kappa_r}{\kappa_{-r}} = \frac{k_r}{k_{-r}}.$$

Thus,

$$F_0(q) = (q-2) \cdot p_{00} + 2 \cdot (q-1) \cdot p_{01} + q \cdot p_{11} = \frac{q \cdot (q-1) \cdot (q-2) \cdot K_r^2}{q \cdot (q-1) \cdot K_r^2 + 2 \cdot q \cdot K_r + 1} + 2 \cdot \frac{q \cdot (q-1) \cdot K_r}{q \cdot (q-1) \cdot K_r^2 + 2 \cdot q \cdot K_r + 1} + \frac{q}{q \cdot (q-1) \cdot K_r^2 + 2 \cdot q \cdot K_r + 1}, \quad (36)$$

and Eq. (32) at stationary conditions can be written as



$$\begin{aligned}
0 &= \lambda_{Lac} \cdot F_0(q+1) \cdot P_0^s(q+1) - k_{Lac} \cdot P_0^s(q) & \text{for } q = 0 \\
0 &= k_{Lac} \cdot P_0^s(q-1) + \lambda_{Lac} \cdot F_0(q+1) \cdot P_0^s(q+1) - (\lambda_{Lac} \cdot F_0(q) + k_{Lac}) \cdot P_0^s(q) & \text{for } q \geq 1
\end{aligned} \quad (37)$$

whose solution can be given as a recursion for  $q \geq 1$

$$\begin{aligned}
\hat{P}_0^s(q) &= \frac{k_{Lac}}{\lambda_{Lac} \cdot F_0(q)} \cdot \hat{P}_0^s(q-1) \\
\hat{P}_0^s(0) &= \left\{ 1 + \sum_{n \geq 1} \prod_{i=1}^n \frac{k_{Lac}}{\lambda_{Lac} \cdot F_0(i)} \right\}^{-1}.
\end{aligned} \quad (38)$$

Having calculated the stationary distribution for  $Lac_T$  we now know the approximate  $\mathcal{O}(\varepsilon)$  stationary distribution for all three species (free operators and  $Lac_T$ ) as well as the distribution for the operators only

$$\hat{P}_0^s(Lac_T, O_{Yfp}, O_{Cfp}) = \hat{P}_0^s(Lac_T) \cdot p_{O_{Yfp} O_{Cfp}}, \quad (39)$$

$$\hat{P}_0^s(O_{Yfp}, O_{Cfp}) = \sum_{Lac_T \geq 0} \hat{P}_0^s(Lac_T) \cdot p_{O_{Yfp} O_{Cfp}}. \quad (40)$$

Note that fixing the total LacI repressor concentration does not eliminate the repressor's extrinsic noise contribution. The reason is that the target promoter dynamics depends on free LacI levels rather than total LacI. A LacI molecule bound to the *cfp* promoter is practically non-existent for the *yfp* promoter and vice versa, implying that free LacI will fluctuate even if total LacI is constant. Elimination of the LacI extrinsic noise contribution would result from assuming that the unbound (free) LacI remains constant.

#### 4. Stochastic simulations

In order to compare the approximation with numerical results we performed simulations of the system using the Gillespie algorithm and the parameter values of Table II. Equation (38) agrees with the distribution obtained from samples taken out of one long Gillespie simulation (compared in Figure 3(a); for the Gillespie simulation sampling was done with constant  $\Delta t = 5 \cdot 10^{-2}$ ). The error norm was calculated as

$$Error = \sqrt{\sum_{Lac_T=0}^{Lac_T_{max}} \left( \hat{P}_0^s(Lac_T) - P_N^{Gillespie}(Lac_T) \right)^2}, \quad (41)$$

which is expected to scale with the inverse square root of the number of samples  $N$  taken from the Gillespie simulation. The simulation results are consistent with this expectation (Figure 3(b)).

Furthermore, since the analytical calculations only refer to promoter states, we performed stochastic simulations to calculate levels of Cfp and Yfp proteins. In Figure 4(a), a scatter plot of the Cfp versus Yfp concentration is shown, obtained from a stochastic simulation of the two promoter model (Eq. (16)). For this simulation, we assume that only one LacI molecule exists, and constant total LacI content is assumed. Evidently, LacI can repress only one operator at a

time, and thus the Cfp and Yfp protein contents are negatively correlated. This negative correlation results in a physically meaningless negative value for the extrinsic noise, namely  $\eta_{ext}^2 = -0.91$ . This situation is of course problematic, since a coefficient of variation should always be non-negative. The negative value obtained is a consequence of the definition of extrinsic noise (Eq. (12)) not being applicable to this case. In panel 4(b) another simulation is shown with 4 LacI molecules. In this case even though there exist enough molecules to keep both promoters repressed at the same time, the extrinsic noise is still negative ( $\eta_{ext}^2 = -1.00$ ). In the following, we will explain this behavior by focusing on the operators and performing analytical calculations for the operator extrinsic noise.

#### 5. Calculation of extrinsic and intrinsic noise

Finally, we will analyze the extrinsic and intrinsic noise for the operators fluctuating between their free and bound

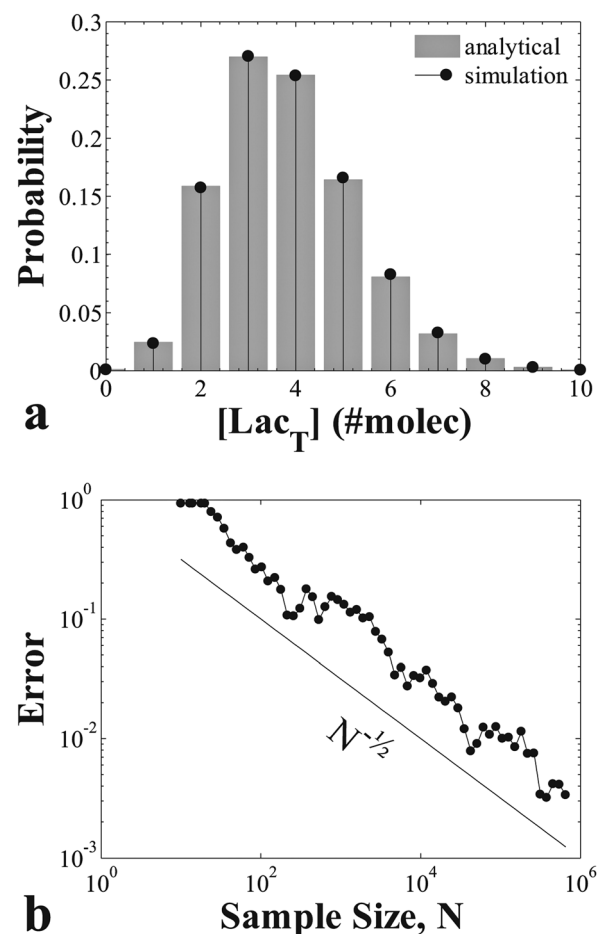


FIG. 3. Comparison of the analytical  $\mathcal{O}(\varepsilon)$  approximation for the stationary probability  $P_0(Lac_T)$  with simulation results. Panel (a): plot of Eq. (38) and of the stationary  $Lac_T$  probability estimate from samples obtained from Gillespie simulations. Panel (b): sampling error falls with the inverse square root of the sample size.

states, using the results from the quasi steady-state approximation. This analysis will provide us with analytical expressions that will give insight into the protein noise since the free operators drive protein expression. Thus, the origin of the unphysical negative extrinsic noise values will be eluci-

dated. Using the expressions for extrinsic and intrinsic noise that were also used by Elowitz et al.,<sup>18</sup> we can calculate the operator noise components given  $\text{Lac}_T$

$$\eta_{\text{oper}|Lac_T}^2 = \frac{\langle (O_{Yfp} - O_{Cfp})^2 \rangle}{2 \cdot \langle O_{Yfp} \rangle \cdot \langle O_{Cfp} \rangle} = \frac{\sum_{i=0}^1 \sum_{j=0}^1 (i-j)^2 \cdot p_{ij}}{2 \cdot \left( \sum_{i=0}^1 i \cdot \sum_{j=0}^1 p_{ij} \right) \cdot \left( \sum_{j=0}^1 j \cdot \sum_{i=0}^1 p_{ij} \right)} \quad (42)$$

$$\eta_{\text{oper}|Lac_T}^2 = \frac{p_{01}}{(p_{01} + p_{11})^2} = \frac{q \cdot K_r}{(q \cdot K_r + 1)^2} \cdot (q \cdot (q-1) \cdot K_r^2 + 2 \cdot q \cdot K_r + 1),$$

$$\eta_{\text{oper}|Lac_T}^2 = \frac{\langle O_{Yfp} \cdot O_{Cfp} \rangle - \langle O_{Yfp} \rangle \cdot \langle O_{Cfp} \rangle}{\langle O_{Yfp} \rangle \cdot \langle O_{Cfp} \rangle} = \frac{\sum_{i=0}^1 \sum_{j=0}^1 i \cdot j \cdot p_{ij} - \left( \sum_{i=0}^1 i \cdot \sum_{j=0}^1 p_{ij} \right) \cdot \left( \sum_{j=0}^1 j \cdot \sum_{i=0}^1 p_{ij} \right)}{\left( \sum_{i=0}^1 i \cdot \sum_{j=0}^1 p_{ij} \right) \cdot \left( \sum_{j=0}^1 j \cdot \sum_{i=0}^1 p_{ij} \right)} \quad (43)$$

$$\eta_{\text{oper}|Lac_T}^2 = \frac{p_{11}}{(p_{01} + p_{11})^2} - 1 = -\frac{q \cdot K_r^2}{(q \cdot K_r + 1)^2},$$

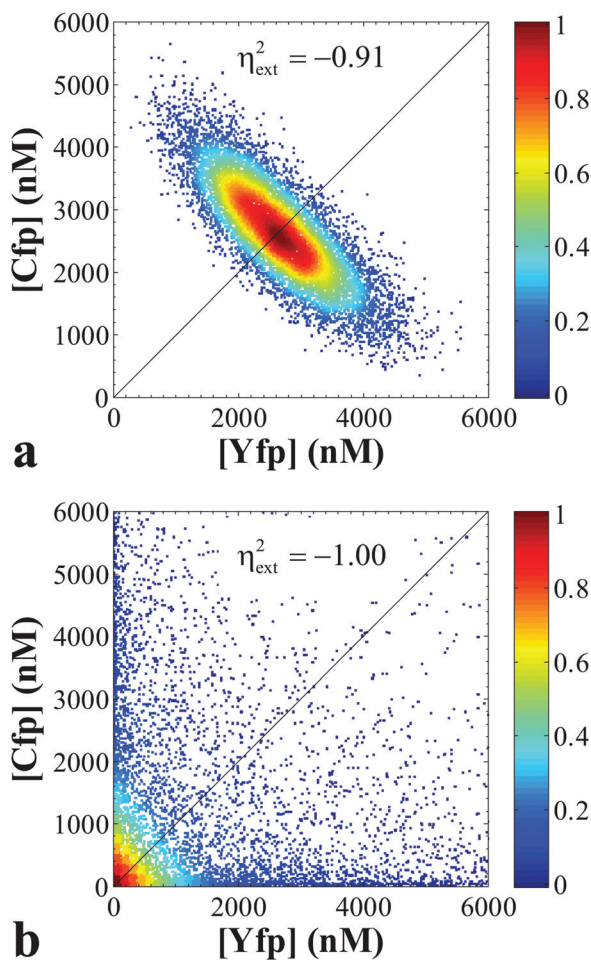


FIG. 4. (Color) Scatter plots of the Yfp versus Cfp concentration assuming constant repressor content ( $k_{Lac} = 0 \text{ nM/min}$ ,  $\lambda_{Lac} = 0 \text{ min}^{-1}$ ). Panel (a): total LacI content equal to 1 molecule; parameter values:  $k_r = 10 \text{ (nM} \cdot \text{min)}^{-1}$ ,  $k_{-r} = 1 \text{ min}^{-1}$ ,  $k_m = 10 \text{ min}^{-1}$ ,  $k_p = 10 \text{ min}^{-1}$ ,  $\lambda_m = 0.4 \text{ min}^{-1}$ ,  $\lambda_p = 0.1 \text{ min}^{-1}$ . Panel (b): total LacI content equal to 4 molecules; parameter values:  $k_r = 50 \text{ (nM} \cdot \text{min)}^{-1}$ ,  $k_{-r} = 1 \text{ min}^{-1}$ ,  $k_m = 10 \text{ min}^{-1}$ ,  $k_p = 1000 \text{ min}^{-1}$ ,  $\lambda_m = 0.4 \text{ min}^{-1}$ ,  $\lambda_p = 0.1 \text{ min}^{-1}$ . Colors indicate relative density of points on the Yfp-Cfp plane: warmer colors correspond to higher densities.

where we have used the symmetry  $p_{01} = p_{10}$ , and  $p_{00}$ ,  $p_{01}$ ,  $p_{11}$  are given by Eqs. (35). Figure 5 presents plots of the intrinsic and extrinsic noise expressions given above. Immediately apparent is the fact that the calculated operator extrinsic noise is *always* negative, which has no physical meaning. This happens because of the negative correlations resulting from the competitive utilization of the common repressor pool. Furthermore, it is interesting to investigate the limiting behavior of the two noise components for high repressor contents ( $q \rightarrow \infty$ ) and for very strong repression ( $K_r \rightarrow \infty$ )

$$\lim_{q \rightarrow \infty} \eta_{\text{oper}|Lac_T}^2 = 0, \quad (44)$$

$$\lim_{K_r \rightarrow \infty} \eta_{\text{oper}|Lac_T}^2 = -\frac{1}{q}.$$

Therefore, the negative correlations just mentioned become progressively negligible for abundant repressor contents. This happens because just a single molecule is required to repress an operator; thus, the change in repressor content brought about by that single molecule becomes negligible is the repressor is abundant. On the other hand, the negative correlations become larger and reach a limiting value for stronger repression. This result is also reasonable, since the tighter the repression, the more probable it is that a repressor molecule will be bound to one operator, thereby depleting the repressor pool.

The above expressions (42) and (43) hold for a fixed value for the  $\text{Lac}_T$  content. In order to calculate the operator noise components for the case where Lac I is produced and degraded (assuming fast operator fluctuations), we have to use the probabilities for the operator states as given by Eq. (40)

$$\eta_{\text{oper}}^2 = \frac{\sum_{Lac_T \geq 0} \hat{P}_0^s(Lac_T) \cdot p_{01}}{\left( \sum_{Lac_T \geq 0} \hat{P}_0^s(Lac_T) \cdot (p_{01} + p_{11}) \right)^2}, \quad (45)$$

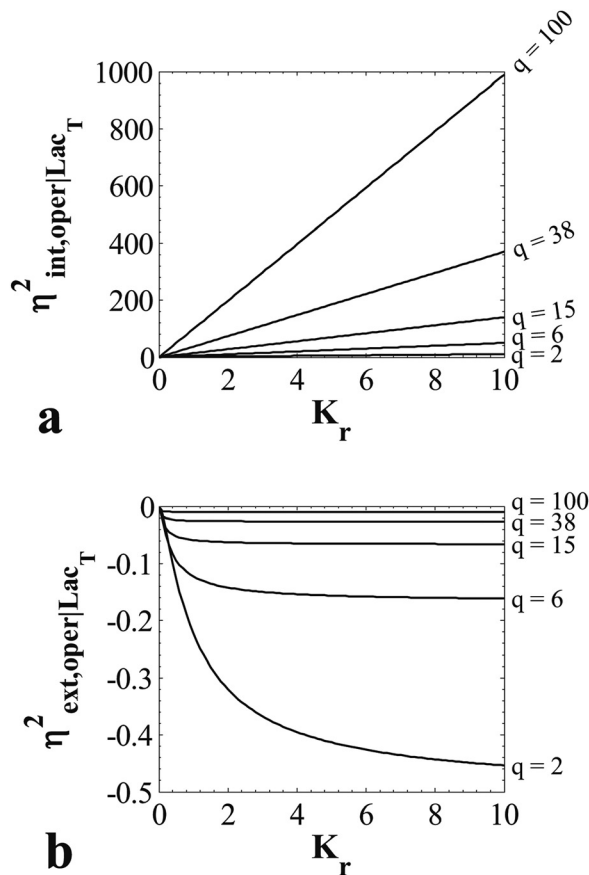


FIG. 5. Plots of the intrinsic (panel a) and extrinsic (panel b) noise for given  $q = \text{Lac}_T$  concentration as calculated from Eqs. (42) and (43).

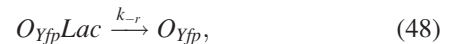
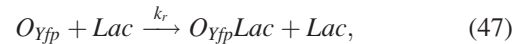
$$\eta_{ext}^2 = \frac{\sum_{\text{Lac}_T \geq 0} \hat{P}_0^s(\text{Lac}_T) \cdot p_{11}}{\left( \sum_{\text{Lac}_T \geq 0} \hat{P}_0^s(\text{Lac}_T) \cdot (p_{01} + p_{11}) \right)^2} - 1. \quad (46)$$

Numerical results indicate that the extrinsic noise as calculated from Eq. (46) is zero for an extended range of the parameter space even though there is clearly an extrinsic noise source due to the LacI fluctuations. Thus, the spurious negative extrinsic effect that we identified analytically for constant LacI content has a counterpart in the case where LacI is allowed to fluctuate.

## B. Non-competitive repressor utilization

Finally, in order to show that indeed it is the competitive repressor utilization that results in the negative feedback

between genes and the paradox of negative extrinsic noise values, we considered the case where genes do not alter the free repressor concentration. Thus, reactions (2) and (3) are substituted with the following:



where  $O_{Yfp}\text{Lac}$  should now be thought of as the operator at an off-state rather than an operator–Lac I complex. Similar reactions can be written for the Cfp. Note that the reaction propensities do not change, but now the genes are not exerting any feedback on the free Lac I concentration. Thus, the Master equation for this system becomes (compare with Eq. (21))

$$\begin{aligned} \frac{\partial \hat{P}}{\partial t} = & (E_{\text{Lac}_T}^{-1} - 1) (k_{\text{Lac}} \cdot \hat{P}) + (E_{\text{Lac}_T} - 1) (\lambda_{\text{Lac}} \cdot \text{Lac}_T \cdot \hat{P}) + \\ & (E_{O_{Yfp}} - 1) (k_r \cdot O_{Yfp} \cdot \text{Lac}_T \cdot \hat{P}) \\ & + (E_{O_{Yfp}}^{-1} - 1) (k_{-r} \cdot (O_{Yfp,T} - O_{Yfp}) \cdot \hat{P}) + \\ & (E_{O_{Cfp}} - 1) (k_r \cdot O_{Cfp} \cdot \text{Lac}_T \cdot \hat{P}) \\ & + (E_{O_{Cfp}}^{-1} - 1) (k_{-r} \cdot (O_{Cfp,T} - O_{Cfp}) \cdot \hat{P}). \end{aligned}$$

Note that in this case the species  $O_{Yfp}\text{Lac}$  merely stands for the Yfp operator at the off state, since free Lac I is not depleted after interacting with the free operator (similarly for the Cfp operator). Thus, now  $\text{Lac}_T$  is always equal to the free

Lac I. If we now apply the summing operator  $\sum_{O_{Yfp}=0}^1 \sum_{O_{Cfp}=0}^1$  to Eq. (49) we get (compare with Eq. (22))

$$\begin{aligned} \frac{\partial \hat{P}(\text{Lac}_T; t)}{\partial t} = & (E_{\text{Lac}_T}^{-1} - 1) (k_{\text{Lac}} \cdot \hat{P}(\text{Lac}_T; t)) \\ & + (E_{\text{Lac}_T} - 1) (\lambda_{\text{Lac}} \cdot \text{Lac}_T \cdot \hat{P}(\text{Lac}_T; t)), \end{aligned}$$

which shows that, in this case,  $\text{Lac}_T$  follows Poisson distribution:

$$\hat{P}^s(\text{Lac}_T; t) = \left( \frac{k_{\text{Lac}}}{\lambda_{\text{Lac}}} \right)^{\text{Lac}_T} \cdot \exp\left(-\frac{k_{\text{Lac}}}{\lambda_{\text{Lac}}}\right) \cdot \frac{1}{\text{Lac}_T!}.$$

Furthermore, by applying the fast operator fluctuations assumption similarly as before we obtain the following linear system and its solution for the probabilities of the operator states (compare with Eq. (34) and the solution (35))

$$\begin{bmatrix} -2 \cdot q \cdot \kappa_r & \kappa_{-r} & \kappa_{-r} & 0 \\ q \cdot \kappa_r & -\kappa_{-r} - q \cdot \kappa_r & 0 & \kappa_{-r} \\ q \cdot \kappa_r & 0 & -\kappa_{-r} - q \cdot \kappa_r & \kappa_{-r} \\ 0 & q \cdot \kappa_r & q \cdot \kappa_r & -2 \cdot \kappa_{-r} \end{bmatrix} \cdot \begin{bmatrix} p_{11} \\ p_{01} \\ p_{10} \\ p_{00} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (52)$$

$$\begin{aligned}
p_{00} &= \frac{q^2 \cdot K_r^2}{(q \cdot K_r + 1)^2}, \\
p_{01} &= p_{10} = \frac{q \cdot K_r}{(q \cdot K_r + 1)^2}, \\
p_{11} &= \frac{1}{(q \cdot K_r + 1)^2}.
\end{aligned} \quad (53)$$

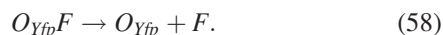
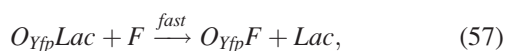
Now, the intrinsic and extrinsic noise values, given the total Lac I, are calculated to be

$$\eta_{oper|Lac_T}^2 = \frac{p_{01}}{(p_{01} + p_{11})^2} = q \cdot K_r, \quad (54)$$

$$\eta_{ext}^2 = \frac{p_{11}}{(p_{01} + p_{11})^2} - 1 = 0. \quad (55)$$

Evidently, the extrinsic noise is always zero in this case due to the fact that the free repressor concentration remains constant.

This analysis shows that the problem of negative correlations resulting from a common pool of a regulatory molecule could be circumvented by the following regulatory strategy. Assume that one could engineer a system in which the regulatory molecule, through a rapid interaction with the operator, induces a reversible change in the state of the latter. After this process, the operator is in the off-state in which it remains for some time, and the regulatory molecule is free to induce the same change to the other operator. Thus, the regulatory molecules are not occupied and the aforementioned negative correlations will no longer represent a problem. Note that the change from the on- to the off-state of the operator can involve a third species, provided that the latter exists at high copy numbers in the cell. If we denote this species by F, the pertinent mechanism could be



Similar reactions hold for Cfp. Since the second step is required to be fast,  $O_{Yfp}Lac$  is a short-lived intermediate that exists in low concentrations, and thus competitive utilization of the Lac species is not an issue. Further, this hypothetical system requires that F exists in high copy numbers, thus, even though a common pool of F exists, the abundance of this species results in negligible competitive utilization when the  $O_{Yfp}Lac$  complex is formed. A natural or engineered system that exhibits such a behavior could entail histone modifications of +1 nucleosomes in eukaryotic promoters.<sup>45</sup> More specifically, histone acetyltransferases or histone deacetyltransferases could modify the accessibility of promoters to transcription factors as well as their transcriptional status, without having to occupy a binding site.

#### IV. SUMMARY

It is widely accepted that noise in gene networks can be decomposed into extrinsic and intrinsic contributions. How-

ever, this is only valid under the premise that genes do not exert any feedback on their inputs; thus, they can neither modulate the extrinsic noise nor interact with each other through a common input.

Here, we have shown that this assumption falls short even in a simple case in which two genes share a common repressor pool. Starting from a simplified reaction network for the two promoter system invented by Elowitz *et al.*<sup>18</sup> to quantify the contributions of extrinsic versus intrinsic noise, we focused on the operator states and calculated the corresponding noise values. We found that due to the competitive repressor utilization, the extrinsic noise value will always be negative if the total repressor is assumed constant. This unrealistic value is due to the negative correlations that result from the fact that a LacI molecule bound to the *yfp* operator, is unavailable for repressing the other operator (*cfp*). If production and depletion of repressor is accounted for (no assumption of constant repressor contents), these negative correlations result in an apparent underestimation of the extrinsic noise. Furthermore, similar conclusions can be reached by analyzing networks with activator species. Competitive utilization of these species by different genes will always result in gene interaction through the common pools and, consequently, in negative correlations in the expression levels of these genes.

Our analysis shows that these effects become less pronounced and finally disappear in the limit of abundant repressor contents (infinitely many molecules). Hence, the presumption that noise can be partitioned into extrinsic and intrinsic components is subject to limitations. In particular, the species' concentrations that act as inputs to the genes of interest must be high enough so that depletion effects are negligible; otherwise, the calculated (or experimentally measured) extrinsic and intrinsic noise values may not be meaningful.

What may be a viable alternative to the intrinsic and extrinsic noise conundrum described above? One option is to rely on information theory to define intrinsic and extrinsic components of stochasticity. Specifically, with sufficient data points (as it is usually the case for flow cytometry data) the mutual information  $I(c,y)$  between the two reporters could be used to estimate extrinsic contributions to gene expression fluctuations. Then, one can calculate the normalized intrinsic and extrinsic entropy factors as follows:

$$F_{int} = \frac{H(y|c) + H(c|y)}{H(c) + H(y)}, \quad (59)$$

$$F_{ext} = \frac{2I(c,y)}{H(c) + H(y)}, \quad (60)$$

where H denotes the Shannon entropy. These factors ( $F_{int}$  and  $F_{ext}$ ) will be always positive, will add up to 1, and may serve as robust descriptors of intrinsic and extrinsic contributions to gene expression noise.

While the drastic effects of competitive regulator utilization described above have not been observed in microbial systems so far,<sup>18,35</sup> further studies on different promoters or engineered systems may reveal the indeterminacy of extrinsic noise. This view is supported by the fact that the



concentration of several transcription factors is very low within the cell, for instance, the copy number of the LacI repressor is roughly 10 in an *E. coli* cell.<sup>24</sup> Moreover, these effects are likely to be more prominent in mammalian systems due to the complexity of gene regulation in higher eukaryotes, where complexes of many proteins (some of which may be in low abundance) must be assembled for transcription initiation. Indeed, scatter plots from two-color reporter assays in at least two recent studies suggest that the promoters driving the reporter genes may be competing for extrinsic regulators.<sup>33,34</sup> Downstream feedback loops may further amplify such competition effects.

Taken together, the above results and these recent studies indicate that gene expression may occasionally be reliant on limited intracellular resources. In such cases, negative regulatory effects stemming from competition for limited regulator pools alone may constitute mechanisms for gene regulation in addition to the biomolecular interactions between protein and DNA species. This emphasizes the integrated nature of all intracellular processes, and the fact that the expression of individual genes is embedded in a complex biochemical environment with implications that are still insufficiently understood.

## ACKNOWLEDGMENTS

This work was supported by the NIH Director's New Innovator Award Program [Grant No: 1DP2 OD006481-01] and by NSF grant BIO IOS (Integrative Organismal Systems) 1021675. The authors would also like to thank the anonymous reviewers for their constructive comments and suggestions.

- <sup>1</sup>H. C. Berg, *Random Walks in Biology* (Princeton University, Princeton, NJ, 1993).
- <sup>2</sup>A. Einstein, *Ann. Phys.* **17**, 549 (1905).
- <sup>3</sup>A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, UK, 2003).
- <sup>4</sup>S. H. Strogatz, *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering* (Westview Press, Cambridge, MA, USA, 2001).
- <sup>5</sup>V. S. Anishchenko, V. Astakhov, A. Neiman, T. Vadivasova, and L. Schimansky-Geier, *Nonlinear Dynamics of Chaotic and Stochastic Systems: Tutorial and Modern Developments*, 2nd ed. (Springer, New York, 2007).
- <sup>6</sup>A.-L. Barabási, *Linked: How Everything Is Connected to Everything Else and What It Means* (Plume, New York, NY, USA, 2003).
- <sup>7</sup>A. L. Barabási and Z. N. Oltvai, *Nat. Rev. Genet.* **5**(2), 101 (2004).
- <sup>8</sup>K. Wiesenfeld and F. Moss, *Nature (London)* **373**(6509), 33 (1995).
- <sup>9</sup>G. Balázsi, A. van Oudenaarden, and J. J. Collins, *Cell* **144**(6), 910 (2011).
- <sup>10</sup>A. Eldar and M. B. Elowitz, *Nature (London)* **467**(7312), 167 (2010).
- <sup>11</sup>M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, *Nat. Rev. Genet.* **6**(6), 451 (2005).

- <sup>12</sup>N. Maheshri and E. K. O'Shea, *Annu. Rev. Biophys. Biomol. Struct.* **36**, 413 (2007).
- <sup>13</sup>A. Raj and A. van Oudenaarden, *Cell* **135**(2), 216 (2008).
- <sup>14</sup>C. V. Rao, D. M. Wolf, and A. P. Arkin, *Nature (London)* **420**(6912), 231 (2002).
- <sup>15</sup>K. F. Murphy, G. Balázsi, and J. J. Collins, *Proc. Natl Acad. Sci. U.S.A.* **104**(31), 12726 (2007).
- <sup>16</sup>K. F. Murphy, R. M. Adams, X. Wang, G. Balázsi, and J. J. Collins, *Nucleic. Acids Res.* **38**(8), 2712 (2010).
- <sup>17</sup>P. S. Swain, M. B. Elowitz, and E. D. Siggia, *Proc. Natl Acad. Sci. U.S.A.* **99**(20), 12795 (2002).
- <sup>18</sup>M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Science* **297**(5584), 1183 (2002).
- <sup>19</sup>M. Thattai and A. van Oudenaarden, *Proc. Natl. Acad. Sci. U. S. A.* **98**(15), 8614 (2001).
- <sup>20</sup>J. J. Tyson and H. G. Othmer, *Prog. Theor. Biol.*, **5**, 1 (1978).
- <sup>21</sup>C. Tan, P. Marguet, and L. C. You, *Nat. Chem. Biol.*, **5**(11), 842 (2009).
- <sup>22</sup>F. J. Isaacs, J. Hasty, C. R. Cantor, and J. J. Collins, *Proc. Natl Acad. Sci. U.S.A.* **100**(13), 7714 (2003).
- <sup>23</sup>P. Guptasarma, *BioEssays* **17**(11), 987 (1995).
- <sup>24</sup>D. Normanno, F. Vanzi, and F. S. Pavone, *Nucleic Acids Res.* **36**(8), 2505 (2008).
- <sup>25</sup>W. Bialek, F. Rieke, R. R. de Ruyter van Steveninck, and D. Warland, *Science* **252**(5014), 1854 (1991).
- <sup>26</sup>J. K. Douglass et al., *Nature (London)* **365**(6444), 337 (1993).
- <sup>27</sup>Z. Gingl, L. B. Kiss, and F. Moss, *Europhys. Lett.* **29**(3), 191 (1995).
- <sup>28</sup>K. Wiesenfeld, D. Pierson, E. Pantazelou, C. Dames, and F. Moss, *Phys. Rev. Lett.* **72**(14), 2125 (1994).
- <sup>29</sup>S. M. Bezrukov and I. Vodyanoy, *Nature (London)* **378**(6555), 362 (1995).
- <sup>30</sup>P. C. Gailey, A. Neiman, J. J. Collins, and F. M. Moss, *Phys. Rev. Lett.* **79**(23), 4701 (1997).
- <sup>31</sup>J. J. Collins, C. C. Chow, and T. T. Imhoff, *Nature (London)* **376**(6537), 236 (1995).
- <sup>32</sup>E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, *Nat. Genet.* **31**(1), 69 (2002).
- <sup>33</sup>T. M. Neild-Nguyen, A. Parisot, C. Vignal, P. Rameau, D. Stockholm, J. Picot, V. Allo, C. Le Bec, C. Laplace, and A. Paldi, *Differentiation* **76**(1), 33 (2008).
- <sup>34</sup>A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, *PLoS Biol.* **4**(10), e309 (2006).
- <sup>35</sup>J. M. Raser and E. K. O'Shea, *Science* **304**(5678), 1811 (2004).
- <sup>36</sup>A. Becskei, B. B. Kaufmann, and A. van Oudenaarden, *Nat. Genet.* **37**(9), 937 (2005).
- <sup>37</sup>D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty, *Nature (London)* **439**(7078), 861 (2006).
- <sup>38</sup>W. J. Blake, G. Balázsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, and J. J. Collins, *Mol. Cell* **24**(6), 853 (2006).
- <sup>39</sup>W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins, *Nature* **422**(6932), 633 (2003).
- <sup>40</sup>M. Stamatakis and N. V. Mantzaris, *Biophys. J.* **96**(3), 887 (2009).
- <sup>41</sup>D. T. Gillespie, *J. Comput. Phys.* **22**(4), 403 (1976).
- <sup>42</sup>D. T. Gillespie, *J. Phys. Chem.* **81**(25), 2340 (1977).
- <sup>43</sup>N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland-Personal-Library, New York, Amsterdam, 1992).
- <sup>44</sup>C. V. Rao and A. P. Arkin, *J. Chem. Phys.* **118**(11), 4999 (2003).
- <sup>45</sup>C. Z. Jiang and B. F. Pugh, *Nat. Rev. Genet.* **10**(3), 161 (2009).